Interactive comment on "Implications of water management representations for watershed hydrologic modeling in the Yakima River Basin" *by* Jiali Qiu et al.

Anonymous Referee #4

The present work assesses the impact of (i) reservoir management and (ii) agricultural-water withdrawn on the hydrologic behavior of the Yakima River Basin (YRB). The Authors provide a convincing and well documented motivation supporting the present study, due to the (not so well explored) role that reservoir management and agricultural activities could have on the hydrological dynamics. I think that the paper is worth for publication after some minor revisions.

Comment 1 The Authors consider 5 different scenario: R0 – the SWAT model is used to simulate the basins dynamics and neither the reservoirs management operations or the agricultural activities are included; R1 - the SWAT model is extended to consider reser-

voirs management operations; R2 – the reservoirs management practice are modelled with RiverWare (which provide the flowrate downstream the each simulated reservoir according to with a set of management rules) which is then combined with SWAT; R2S2 – leverage on R2 including agricultural activities under the hypothesis that all the demand water comes from reservoirs and streams; R2S1 – is the counterpart of R2S2 in which the agricultural demand is satisfied by superficial aquifers. For each scenario (I would rather say modelling scenario/choice) there is a calibration period and a validation period. Model performance metrics are the correlation coefficient r (note that in the figure is referred as R, please change it) and the Sutcliffe efficiency coefficient ENS. The Authors found a better model performance in the following order R2S1 > R2S2 > R2 > R1 > R0, leading to the conclusion that using RiverWare to model the reservoirs management and modelling the agricultural demands as satisfied by surface bodies is the best option. How does the Authors ensure that ensuing models ordering is not influenced by the fact that a better fit between model results and data (which is at the base of r and ENS metrics) is not an artefact of model flexibility/complexity (i.e., more parameters) rather than being a 'true' more realistic model? See e.g., "A primer for model selection: the decisive role of model complexity" by Hoge et al., 2018 (https://doi.org/10.1002/2017WR021902)

Comment 2 On top of the GSA results the Authors conclude that (line 187-189) : "In general, selected parameters demonstrated similar sensitives among all scenarios, particularly for the ten most sensitive parameters, indicating that the five scenarios captured critical processes regulating water cycling in the basin". I would disagree with the second part of the sentence. The fact that in all the 5 scenarios the top 10 most influential parameters are quite the same suggests a similarity in the behaviour/dynamics/functioning of the 5 models investigated (since for each of them the key parameters are the same), which does not ensure the fidelity with the true-world dynamics! Furthermore, this result seems to contradict in part the relevance of having different modelling strategies for the reservoirs management and/or the agricultural activities, since the 5 model results are mainly ruled by the adopted representation of

common processes, like the moisture conditions (parameter CN2 is always the most important one) or snow melt process/parameters (see Line 190, with 5 parameters occupying the 2nd -6th places). As a matter of fact, the joint inspection of Fig. 3-6 does not reveal a dramatic change in the r and Ens values. It is also true that starting from the 7th position parameters associated with reservoir management and agricultural activities alternate as importance. I would appreciate a more detailed description of the GSA methodology (how it is possible to vary a parameter, and so test its sensitivity on the model outputs, related with the reservoirs, e.g., the RES_K, if the reservoirs are not included in R0? The same for agricultural activities related parameters). Furthermore, is the employed GSA able to provide a quantification of the global sensitivity of each parameter and not just their ranking (e.g., could be that CN2 is dominating the process and the others parameters induce just very small variations in the output)

Comment 3 Caption of Fig. 2, note that should be ' baseline simulation does NOT consider management activities'.

Comment 4 Line 279. "Management schemes developed and evaluated in this study will be transferable and applicable to future SWAT and other watershed models applications for investigating water cycling that is influenced by reservoir operations and water withdrawal for irrigation across broader spatial scales" Which are the developed schemes? It was my understanding that the Authors employed SWAT and RiverWare without any modification to them.

Comment 5 Eq. (1): Vnet seems to me like the water volume after one day, being Vstored the water stored at the beginning of the day. Please clarify that Vflowout is the focus of the diverse management schemes. Eq. (4): Vflowout has been already used to indicate a volume in Eq. (1), please modify the notation in order to avoid confusion.

Comment 6 Line 207: delete water after release.

Comment 7 Line 321-323: "The irrigation operation scheme that used surface water as the single source may have introduced uncertainties to streamflow simulations, since

ground water is also an important water source for irrigation, particularly in dry years in the YRB". The choice of using surface water as the single source for the irrigation surely introduce (increasing, by e.g., lack of knowledge on some parameter, or reducing, e.g., by inserting salient dynamics in the model) uncertainties to streamflow, but I think that in the context of the sentence this choice has to be seen as a 'bias' to the streamflow simulations, i.e., streamflow are biased by having choice to consider only surface water bodies.

Comment 8 Caption of Fig. 6: is it R2S1 based on SWAT or RiverWare?